14

Engineered Proteins in Materials Research

David A. Tirrell, *June* G. *Tirrell,*
Thomas *L.* Mason, *and Maurille J. Fournier*

Peptides and proteins have attracted scientific and technological interest largely because of their intriguing properties as catalysts, receptors, signalling molecules, and therapeutic agents. In attempts to understand and exploit these properties, protein engineering has been used primarily to obtain precious proteins in increased quantities, or to explore systematic alterations in protein sequence through site-directed mutagenesis. Design of protein structures de ***novo*** ("from scratch") has attracted less attention, and has been directed in the main toward studies of protein folding (Kamtekar et al., 1993). Such studies represent a key element in the current vigorous investigation of the connections between amino acid sequence **and** the three-dimensional structures of isolated protein chains in aqueous solution. This chapter describes protein engineering of quite another sort, in which the ***proteinacious*** nature of the product is less important than its ***macromolecular*** character.

Engineered Proteins

The chemical processes used to prepare the macromolecular materials (polymers) of commerce produce complex mixtures of products, in which chain length, sequence, and stereochemistry are characterized by broad statistical distributions. As an example, Figure 14-1 shows a matrix-assisted laser desorption mass spectrum of poly(methyl methacrylate) (PMMA), an important polymer sold under tradenames such as Lucite or Plexiglas. The sample represented in Figure 14-1 is far better defined in terms of molecular weight distribution than is the typical commercial PMMA, yet it is clear from the spectrum that no single molecular species constitutes more **than** a small percentage of the chain population. Similar complexity characterizes the distributions of ***monomer sequences*** in commercial copolymers, and the ***stereochemistry*** of vinyl polymers is generally discussed only in statistical terms. This complexity of structure has had profound consequences for polymer science and technology, because it means that this is a science and technology of ***mixtures.*** This is not necessarily an impediment to the development of useful properties, and conventional polymers are enormously important as practical, commercial materials. At the same time, because mixtures behave

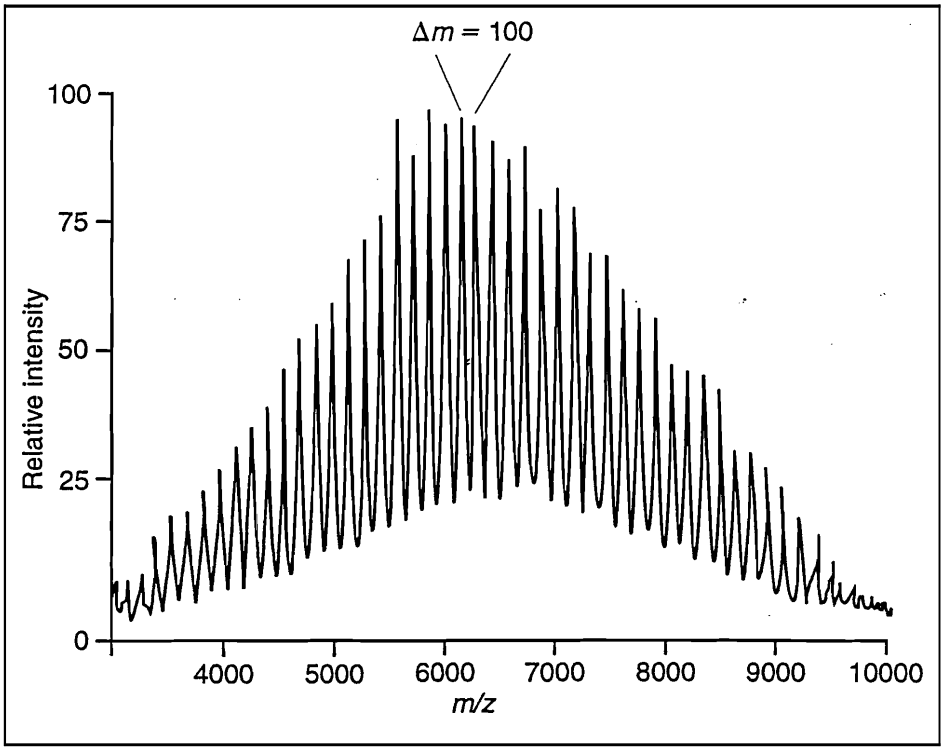


Figure 14-1. MALDI mass spectrum of poly(methyl methacrylate). Reprinted from Bahr et al., 1992, with permission.

differently from pure substances, one wonders what kinds of new materials properties might be achievable through better control of macromolecular architecture.

Protein engineering is now the most powerful means of controlling the architecture of macromolecular substances, and offers an important alternative to conventional polymerization processes as a route to new classes of polymeric materials. The exploitation of protein engineering to create new materials is outlined in schematic form in Figure 14-2. The process begins with design of a target structure, or more typically, a family of related target structures. The targets here are artificial proteins, amino acid copolymers designed to exhibit interesting ***materials*** properties, for example, interesting properties in the solid state, at interfaces, or in liquid crystal phases. The design of such proteins must take into account both "biological" and "materials" concepts in order to succeed. Once designed, the primary sequence of the target artificial protein is encoded into a complementary DNA, which is then made by solid phase synthesis and enzymatic ligation of oligonucleotide fragments. Conversion of the artificial gene to the artifical protein is then accomplished via the same kinds of genetic manipulations and biochemical engineering methods used to produce other classes of heterologous proteins in microbial hosts (Sambrook et al., 1989).

In our own laboratories, we have used this approach to address four issues in materials design and synthesis: (1) design and preparation of macromolecular crystals of

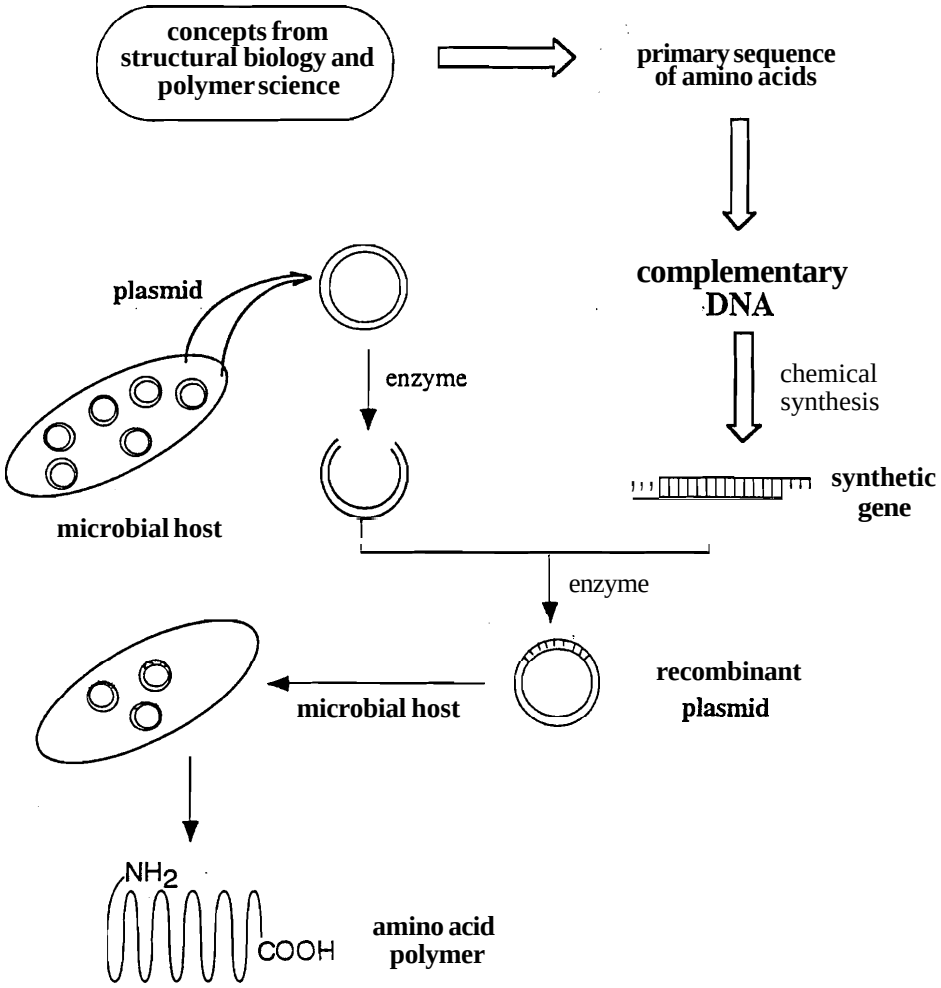


Figure 14-2. Outline of gene construction and protein synthesis.

controlled thickness and surface chemistry, (2) incorporation of nonnatural amino acids, (3) synthesis of rodlike helical macromolecules of uniform chain length, and (4) development of hybrid proteins comprising natural and artificial domains. We describe in this chapter the progress that has been made toward these objectives, and we hope thereby to illustrate some of the exciting opportunities that we perceive at this new interface of materials science and the biological sciences.

Crystal Design: The Solid-State Protein Folding Problem

A logical starting point for the design of macromolecular crystals is provided by the sequence-dependent secondary structures (helices, β-strands, and reverse turns) found in fibrous and globular proteins. For example, a repetitive polypeptide comprising β-strands followed by reverse turns would be expected to assemble into a lamellar ag-

gregate like that shown schematically in Figure 14-3, in which lamellar thickness is controlled by the length of the strands, and surface chemistry is determined by functional groups contributed by amino acid residues confined to the turns. Such lamellar crystals are well known in polymer materials science (Keller, 1957), but are usually only metastable structures trapped by the kinetics of crystal growth. In the design shown in Figure 14-3, the lamellar structure should not only be stable, but should be predictable on the basis of the secondary structural elements used for its construction.

In order to test this approach to crystal design, we have prepared and expressed a family of artificial genes encoding polypeptides built from repeating units represented as sequence **1** (Deguchi et al., 1994; Krejchi et al., 1994,.1996).

$$\{\text{-(AlaGly)}_n\text{ZGly-}\}_x$$

1 $\quad n = 3$ to 6

Z = Ala, Asn, Asp, Glu, Leu, Met, Phe, Ser, Tyr, Val, and ProGlu

In the crystal structures of synthetic aliphatic polyamides (nylons), kinetic factors limit the length of the crystal "stem" to that defined by six to eight lateral hydrogen bonds (Atkins et al., 1992; Dreyfuss & Keller, 1970a, b); the choice of $n = 3$ to 6 in **1** reflects that consideration. It has also been observed that the egg-stalk protein of *Chrysopa flava* (Geddes et al., 1968) folds at intervals of eight amino acid residues, giving rise to a lamellar thickness of about 3 nm. Bulky and polar amino acids in the Z position were chosen because such amino acids should be excluded from the interior and should therefore decorate the surfaces of the lamellar crystal. Glutamic acid in particular was chosen because it is the poorest β-sheet former of the twenty natural amino acids according to the Chou-Fasman predictions of protein structure and conformation (Chou & Fasman, 1974a, b).

Synthesis

The following paragraphs provide a brief description of the methods used in our laboratory to prepare polymers of repeating units such as **1** (Fig. 14-2).

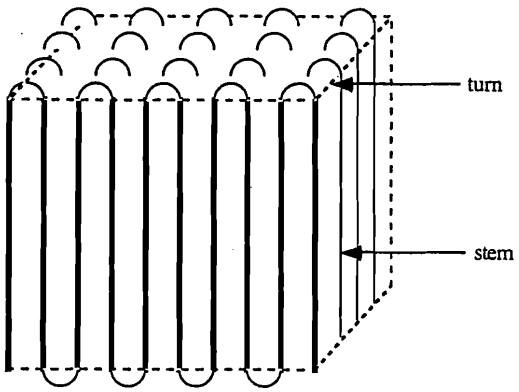


Figure 14-3. Schematic drawing of a chain-folded lamellar crystal. Reprinted from Parkhe et al., 1993, with permission.

```
                                  Stop Gly Ala Gly Ala Gly Ala Gly Ala Gly Pro Glu Gly Ala Gly Ala Gly Ala Gly Pro Glu Gly Ala
AATTCG TAA GGT GCC GGC GCT GGT GCT GGG GCC GGT CCG GAA GGT GCA GGC GCT GGC GCG GGC GCG GGC CCG GAA GGT GCC G
      GC ATT CCA CGG CCG CGA CCA CGA CCC CGG CCA GGC CTT CCA CGT CCG CGA CCG CGC CCG CGC CCG GGC CTT CCA CGG CCTAG
   EcoRI          BanI                                                                  ApaI          BanI   BamHI
```

2

Oligonucleotides encoding one or two repeats of the target repeating unit sequence are prepared via solid phase organic synthesis (McBride & Caruthers, 1983). After purification, the oligonucleotides are phosphorylated at the 5′-termini, annealed, and then ligated into appropriate restriction sites in a bacterial cloning vector. Oligonucleotides are typically designed according to the following rationale. First, as shown in fragment 2 (which encodes two copies of repeating unit **1** with $n = 4$ and Z = ProGlu), a stop codon immediately following the 5' restriction site ensures disruption of the β-galactosidase a fragment encoded in common cloning vectors, and thus allows facile color screening for plasmids carrying the insert. Second, the oligonucleotides are designed so that the sequence encoding the repeating unit is flanked by two restriction sites which are used to isolate the fragments of interest after cloning and amplification (Cappello et al., 1990). Finally, the choice of codons for each amino acid in repeating unit **1** is determined by (1) the pattern of codon use in E. *coli* (Aota et al., 1988), *(2)* avoidance of strict sequence periodicity within the oligonucleotide, **(3)** the need to eliminate all *Ban*I sites except for those flanking the coding sequence, and (4) inclusion of an *Apa*I site to be used for screening transformants containing the insert.

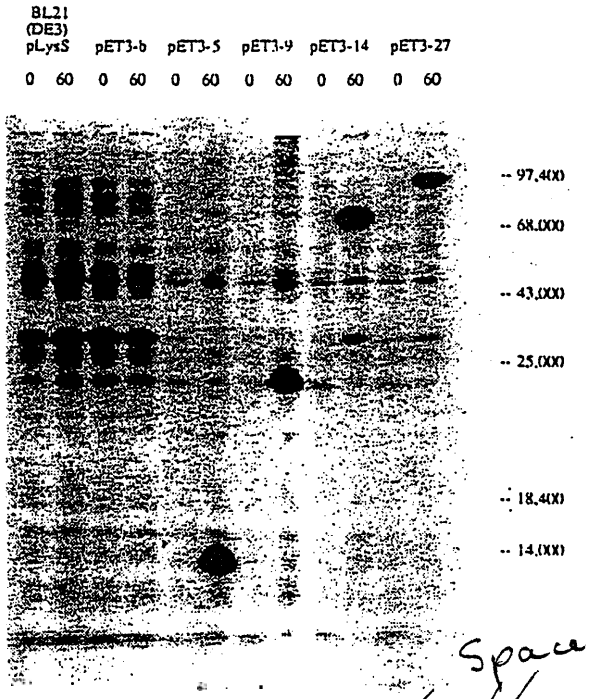


Figure 14-4. In *vivo* labeling of proteins containing $(AlaGly)_3ProGluGly$ repeats in 12% polyacrylamide gels. pET3-5, pET3-9, pET3-14 and pET3-27 contain 5, 9, 14, and 27 repeats of the $(AlaGly)_3ProGluGly$ sequence, respectively. Time points in minutes are relative to IPTG addition. Reprinted from McGrath et al., 1992, with permission.

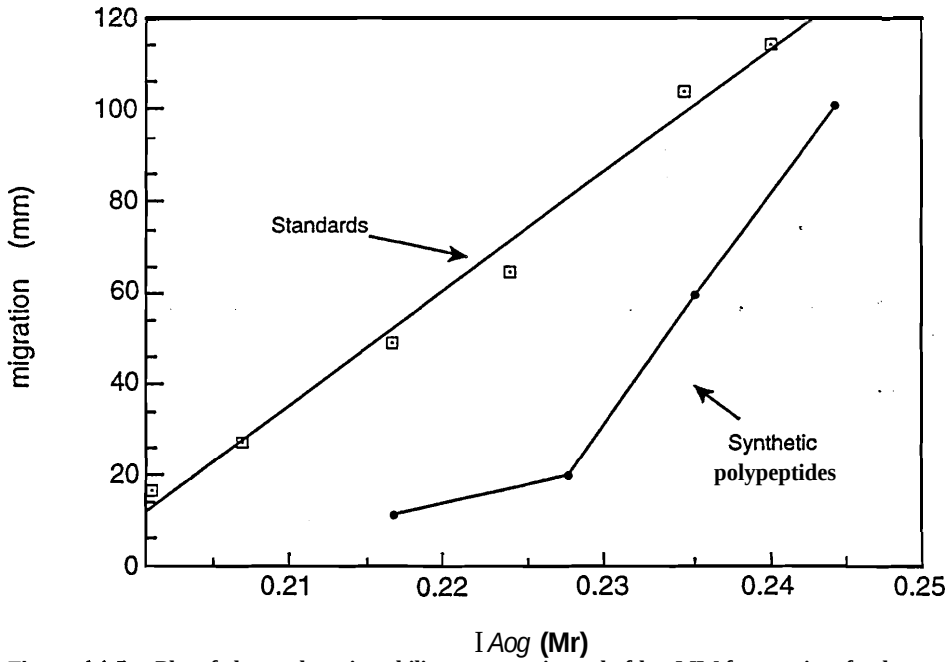


Figure 14-5. Plot of electrophoretic mobility versus reciprocal of log MW for a series of polypeptides containing 5, 9, 14, and 27 repeats of the $(AlaGly)_3ProGluGly$ sequence. Standards: α-lactalbumin, β-lactoglobulin, carbonic anhydrase, egg albumin, bovine serum albumin, phosphorylase B (rabbit muscle). Reprinted from McGrath et al., 1992, with permission.

The presence of the insert in transformed bacterial cells can be confirmed by restriction analysis (for sequence 2, for example, with the restriction enzymes *ApaI* and *BanI*), and standard DNA sequencing methods are used to verify the integrity of the insert. The target DNA fragments are liberated by restriction digestion, purified, and self-ligated to produce a population of multimers. Inversion of repeats of the monomer sequences is suppressed by the fact that nonpalindromic ends are generated in *BanI* digestion of the plasmid. Multimers of appropriate size are inserted (usually after an intermediate cloning step, which allows selection of coding sequences of preferred lengths) into an expression vector, which is chosen to direct bacterial synthesis of the target protein. At this point, the coding sequence may include N- and C-terminal extensions derived from the cloning and expression vectors; these are generally removed later by cleavage with CNBr or with appropriate enzymes.

Target protein synthesis can be demonstrated in a variety of ways. Because artificial proteins such as **1** bind conventional protein stains weakly, radiolabeling protocols or Western blots are most convenient. In a typical labeling experiment, transformed cells are grown in minimal medium supplemented with ^{3}H-glycine and a mixture of unlabeled amino acids lacking glycine. Following induction of protein synthesis, new protein bands corresponding to the target can be detected on autoradiograms of gels (Fig. 14-4). Protein products of sequence **1** may be present either in the soluble fraction of the cell lysate or in inclusion bodies, depending on charge density and the nature of residue(s) Z.

The apparent molecular weights of these artificial proteins as reported by gel electrophoresis are much higher than the expected molecular weights (Fig. 14-5). Highly

acidic polypeptides such as **1** (Z = Glu or ProGlu) probably bind SDS weakly and may, therefore, adopt unusual micellar structures under the conditions of electrophoretic separation. The anomalously low average residue mass of these alanylglycine-rich polymers surely contributes further to this effect.

In our work with proteins such as **1**, neither the artificial genes nor the protein products have proven prohibitively unstable. Plasmids from transformed cells in some cases have been recovered and subjected to electrophoresis; even after 35 generations, no length polymorphism was observed. The protein products are sufficiently stable that they continue to accumulate in transformed cells for 2 to 3 hours (and perhaps longer) following induction (for an exception, see later). Typical yields from batch fermentation procedures have been in the range of 100 mg of protein per liter, and recent experiments using fed-batch fermentation methods have afforded yields of nearly 0.5 g of purified protein per liter of culture.

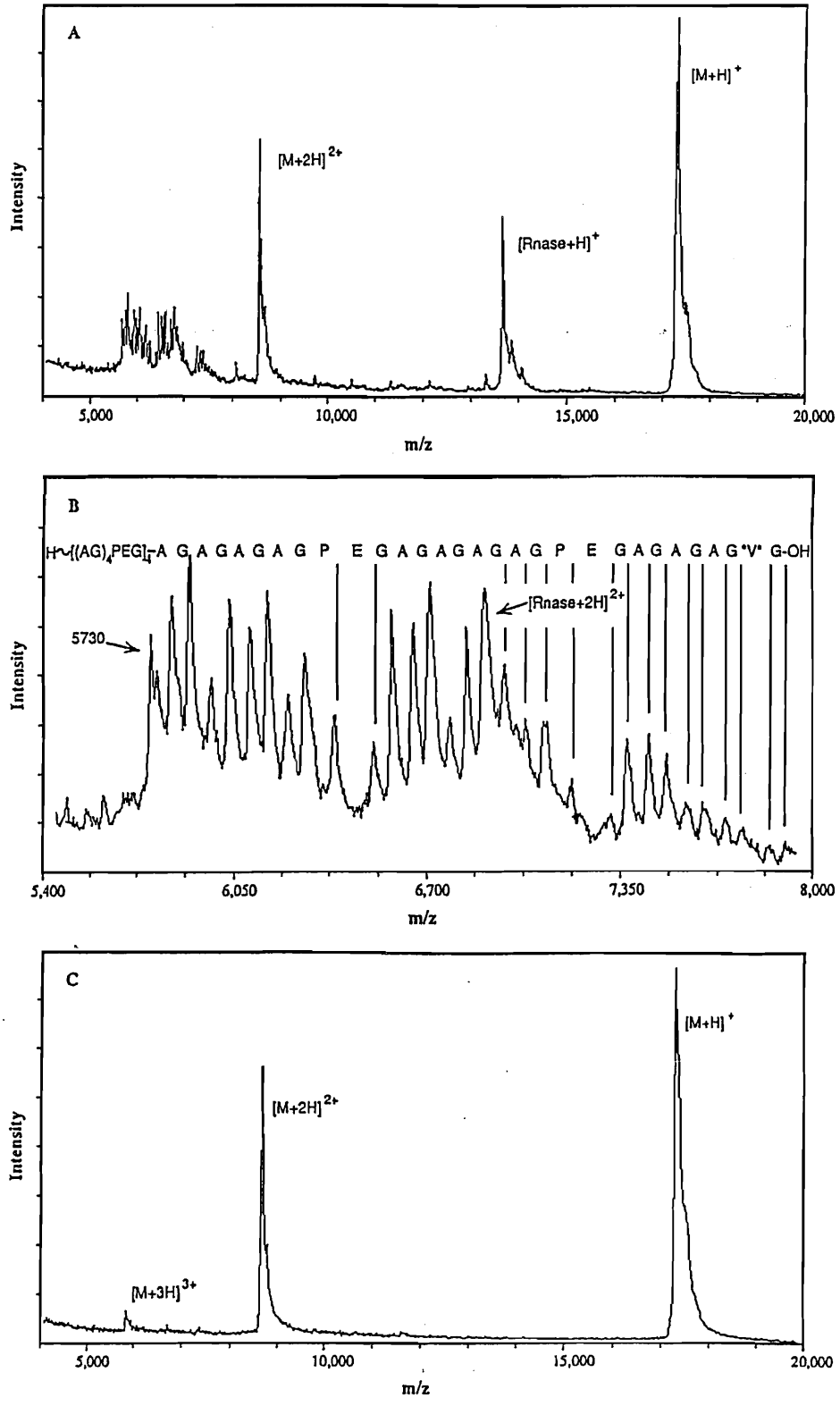
Molecular Structure

The primary structures of polymers of repeating units such as **1** are generally as expected. In one interesting exception, matrix assisted laser desorption (MALDI) mass spectrometry of the protein comprising 14 repeats of the sequence -(AlaGly)$_4$ProGluGly- (3), showed evidence of degradation of the product protein (Beavis et al., 1992) (Fig. 14-6). The spectrum of the isolated protein contained a series of signals that could best be explained by successive additions of amino acid residues, in the N- to the C-terminal direction, consistent with the anticipated amino acid sequence. MALDI mass spectrometry also gives an accurate measure of the mass of the polymer, and shows that mass determinations by SDS polyacrylamide gel electrophoresis are grossly in error.

Solid-State Structure

Proteins built up from repeating units of sequence **1** (Z = ProGlu) do not readily form β-sheets (McGrath et al., 1992). Under most of the crystallization conditions we have investigated, these polymers form amorphous solids, as indicated by wide-angle X-ray powder patterns that consist only of diffuse halos. Moreover, the amide I and amide II bands in the infrared spectrum are observed at 1653 and 1540 cm^{-1}, respectively, and not at the anticipated frequencies (1630 and 1525 cm^{-1}) characteristic of β-structures (Moore & Krimm, 1976). We have considered several explanations including (1) the odd number of amino acids in the repeating unit sequence precludes formation of hydrogen bonds over the full extent of the sheet (McGrath et al., 1992), (2) the geometry of the turn (presumably comprising the ProGlu dyad) is inconsistent with the parallel (or nearly parallel) trajectories required of the flanking strands, and (3) the steric bulk of the ProGlu dyad frustrates packing of sheets at a separation distance consistent with the small size of the Ala and Gly residues comprising the sheets. Whether or

Figure 14-6. Matrix-assisted laser desorption mass spectrometric analysis of {(AlaGly)$_4$ProGluGly}$_{14}$. (A) Spectrum of the target protein. RNAse A is included as an internal standard. (B) Spectrum expanded in the region of low molecular weight contaminants. (C) Mass spectrum of the protein sample after low molecular weight substances were removed by dialysis. Reprinted from Beavis et al., 1992, with permission.


A
[M+H]+
[M+2H]2+
[Rnase+H]+
Intensity
5,000
10,000
15,000
20,000
m/z
B
H~[(AG)4PEG]1-A G A G A G A G P E G A G A G A G A G P E G A G A G A G V G-OH
[Rnase+2H]2+
5730
Intensity
5,400
6,050
6,700
7,350
8,000
m/z
C
[M+H]+
[M+2H]2+
[M+3H]3+
Intensity
5,000
10,000
15,000
20,000
m/z

not any of these explanations is correct, it appears that the regularly folded conformation is destabilized to such an extent that the glassy state is preferred at room temperature or, alternatively, that crystallization is prohibitively slow.

In order to address these concerns, we explored a second generation of designed structures in which Pro was deleted from the repeating unit. We first considered polymers of repeating unit sequence -(AlaGly)$_n$GluGly- (4, n = 3, 4, 5, 6), and found that such polymers form β-sheet structures readily under conditions that failed to produce crystalline forms of the "ProGlu" polymer (Krejchi et al., 1994).

F14-7

The infrared spectrum of {(AlaGly)$_3$GluGly}$_{36}$ (Fig. 14-7) shows amide I, II, and III vibrational modes at 1623, 1521, and 1229 cm^{-1}, respectively, all of which are characteristic of the β-sheet structure (Moore & Krimm, 1976), and a weak amide I vibration at 1698 cm^{-1} indicating the regularly alternating chain direction that defines antiparallel β-sheets (Miyazawa & Blout, 1961). Additional vibrational modes are thought to arise from reverse (β- or y-) turn structures (Krimm & Bandekar, 1986).

Raman spectroscopy of {(AlaGly)$_3$GluGly}$_{36}$ further supports the proposed structure, showing the characteristic amide I band at 1664 cm^{-1} and splitting of the arnide 111 band into two components at 1260 and 1228 cm^{-1} (Frushour & Koenig, 1975;

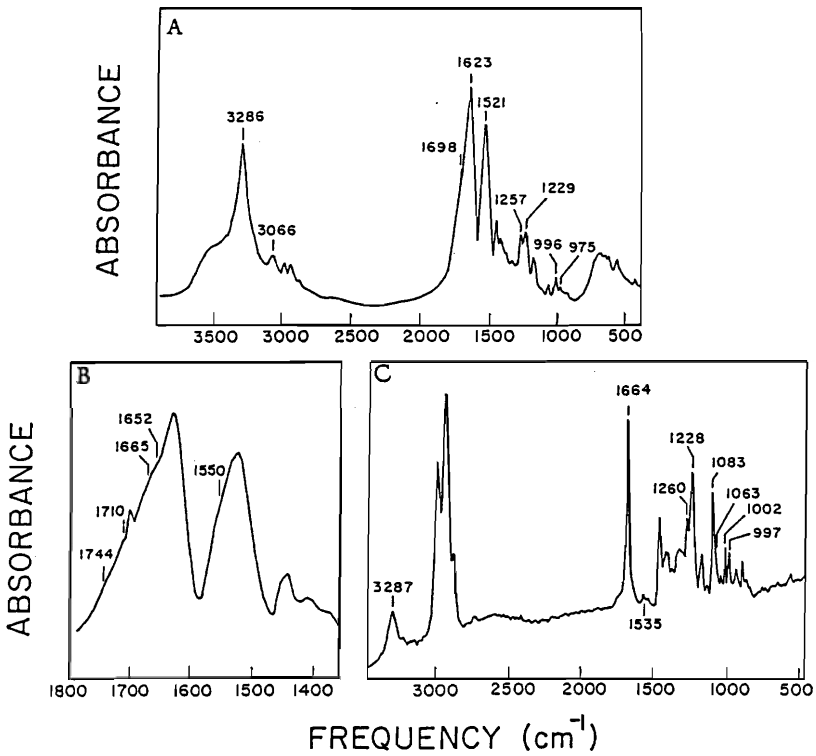


Figure 14-7. Vibrational spectra of {(AlaGly)$_3$GluGly}$_{36}$. (A) Fourier transform infrared spectrum of a KBr pellet of {(AlaGly)$_3$GluGly}$_{36}$ powder. (B) Expansion of the amide I region of the spectrum shown in A. (C) Raman spectrum of {(AlaGly)$_3$GluGly}$_{36}$. Reprinted from Krejchi et al., 1994, with permission.

Table 14-1 Chemical Shifts and Assignments for Selected Peaks Observed in CP/MAS ^{13}C NMR Spectra of $\{(AlaGly)_3GluGly\}_{36}$ and Poly(L-alanylglycine) (β-form)

Chemical Shift (ppm)		
$\{(AlaGly)_3GluGly\}_{36}$	Poly(L-alanylglycine)	Assignment
49.9	48.5	Ala Cα
20.7	20.0	Ala Cβ
171.4	171.8	Ala C=O
43.6	43.3	Gly Cα
171.4	168.4	Gly C=O

Spaces

Moore & Krimm, 1976). Weaker signals in the 1300 to 1330 cm^{-1} region have been attributed to turns (Krimm & Bandekar, 1986).

T14-1

Cross-polarization magic angle spinning nuclear magnetic resonance (CP/MAS NMR) spectra provide additional confirmation that $\{(AlaGly)_3GluGly\}_{36}$ forms antiparallel β-sheets. Table 14-1 shows a comparison between the chemical shifts obtained for poly(L-alanylglycine) (Saito et al., 1984) and for $\{(AlaGly)_3GluGly\}_{36}$. The assignments are consistent with an antiparallel β-sheet structure, though there is evidence for other conformational states as well. For example, a shoulder at 16.8 ppm, assigned to the β-carbon of alanine, arises either from a fraction of the silk I structure (Ishida et al., 1990), or from amino acid residues in turn sequences (as proposed above to explain the infrared and Raman spectra).

T14-2

Strong evidence for a chain-folded lamellar structure as the basic crystalline unit is provided by X-ray diffraction results obtained on the series of polymers, $-\{(AlaGly)_nGluGly\}_x-$ (4) (Krejchi et al., 1997). The dimensions of the corresponding orthorhombic unit cells are listed in Table 14-2, and small angle X-ray scattering gives a long-period spacing of 3.6 nm (for the polymer with n = 3), consistent with the anticipated lamellar thickness.

The unit cell parameters are consistent with those published for silks and for synthetic polypeptides known to adopt β-sheet architectures (Brown & Trotter, 1956; Warwicker, 1960). The value of the unit cell parameter ***a***, 0.948 nm, was assigned on the basis of the second diffraction order spacing at 0.474 nm, which is characteristic of the

Table 14-2 Unit Cell Dimensions for Polymers of $-\{(AlaGly)_nGluGly\}_x-$

Sample	Unit Cell Dimensions (nm)		
	a	*b*	*c*
$\{(AlaGly)_3GluGly\}_{36}$	0.948	1.060	0.695
$\{(AlaGly)_4GluGly\}_{28}$	0.948	1.028	0.695
$\{(AlaGly)_5GluGly\}_{20}$	0.957	0.970	0.695
$\{(AlaGly)_6GluGly\}_{14}$	0.964	0.962	0.695

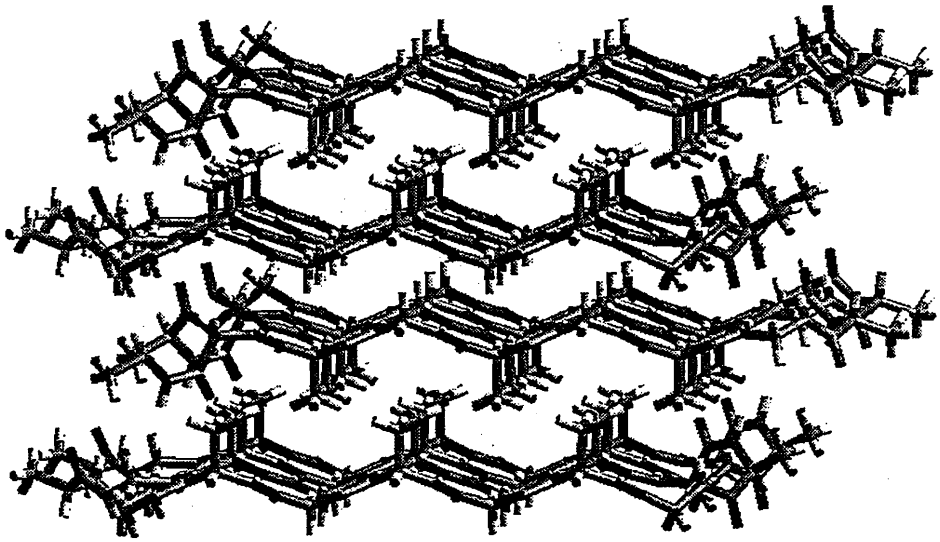


Figure 14-8. Computer-generated representation of the solid-state structure of $\{(AlaGly)_3GluGly\}_{36}$. The bottom portion shows a space-filling representation. Reprinted from Krejchi et al., 1994, with permission.

Not previously published.

interchain distance in hydrogen-bonded, antiparallel /I-sheets (Dreyfuss & Keller, 1970a, b; Hirichsen, 1973; Keller, 1959). Thus a is associated with the hydrogen bond direction. The value 1.060 nm observed for b represents twice the average periodicity of intersheet stacking. The intersheet spacing in β-sheet polypeptides is dependent on amino acid composition (Lucas et al., 1960), since the amino acid side chains protrude from the surfaces of the sheets and must be accommodated in the intersheet volume. The intersheetspacing for polyglycinehas been reported as 0.34 nm (Nemethy & Printz, 1972), while that of *Nephila senegalensis* fibroin, which contains a high percentage of amino acid residues with bulky side chains, is 0.79 nm (Warwicker, 1960). The values reported for poly(L-alanylglycine), 0.44 nm (Fraser et al., 1965), and the β-form of poly(L-alanine), 0.54 nm (Brown & Trotter, 1956), are consistent with the results obtained for the series of polymers 4.

F14-8

Figure 14-8 shows a computer generated representation of the unit cell of polymer 4 (n = 3) in which the β-sheets are stacked along the b axis, the c axis extends horizontally in the plane of the image, and the hydrogen bond direction, a, is perpendicular to the plane of the page. The polymer chains in the crystals must fold back at the lamellar surfaces and reenter the crystalline lamellae, as indicated by the crystal dimensions observed, since the lamellar thickness is always shorter than the molecular length of the chains.

Since $\{(AlaGly)_3GluGly\}_{36}$ exhibits a well-defined β-sheet conformation and a uniform lamellar structure, it serves as a good model for studying the relationships between hydration and crystal structure. Infrared and Raman spectral studies of the $\{(AlaGly)_3GluGly\}_{36}$ polypeptide have shown that hydration is a stepwise process (Chen et al., 1995). At low water contents, the crystalline regions are not affected (Murthy et al., 1989; Tirrell et al., 1979; Vergalati et al., 1993). As the water content increases, the sheets become accessible to water and changes in intersheet spacing, and ultimately chain

conformation, occur. Figure 14-9 shows the effects of humidity on the infrared spectrum of the carboxylate form of the sample. Increasingly marked effects on both the COO– and the amide I and II signals are observed as the water content increases. Figure 14-10 shows the intersheet distance as a function of humidity as determined by wide angle X-ray scattering, infrared spectra of the amide I and II regions at different relative humidi-

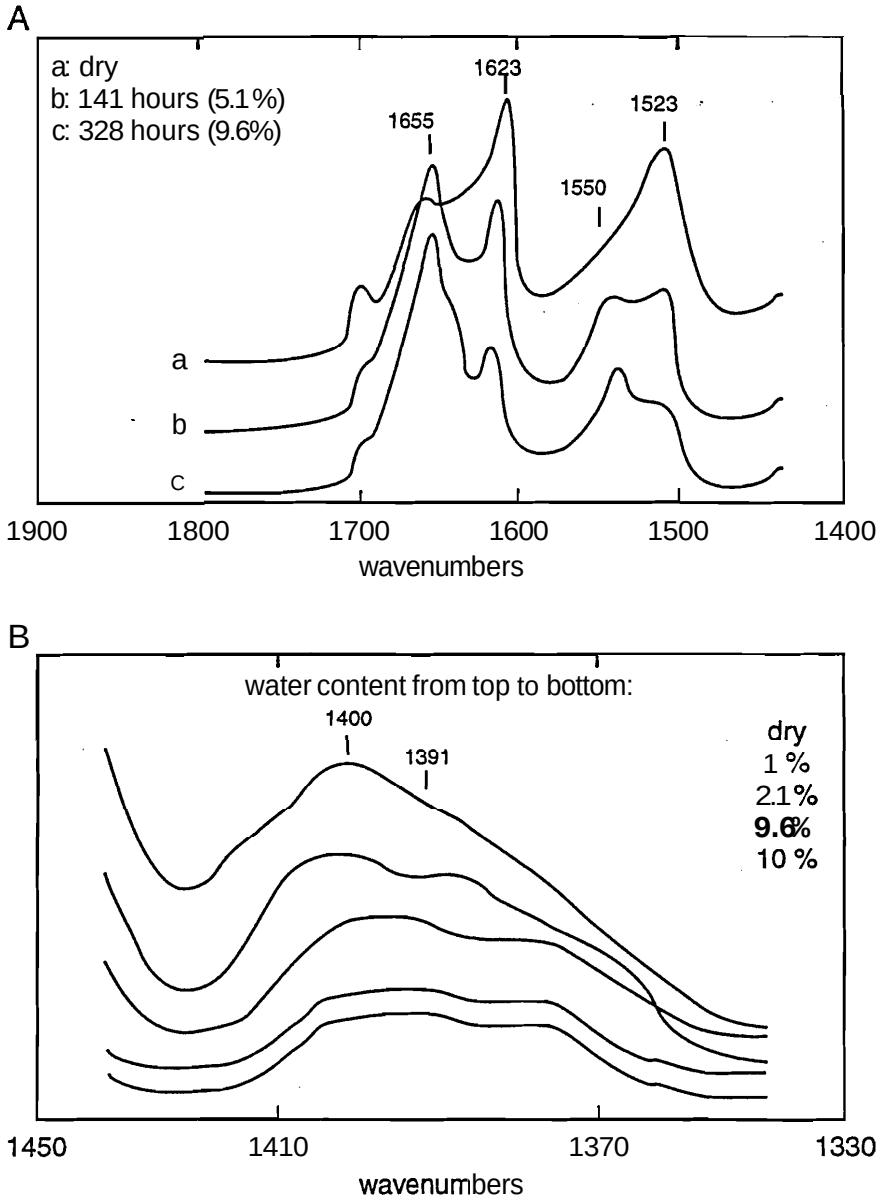


Figure 14-9. (A) Infrared spectra of the amide I and II region of $\{(AlaGly)_3GluGly\}_{36}$ in the carboxylate form as a function of hydration. (B) Effect of hydration on the symmetric COO– band in the infrared spectrum. Reprinted from Chen et al., 1995, with permission.

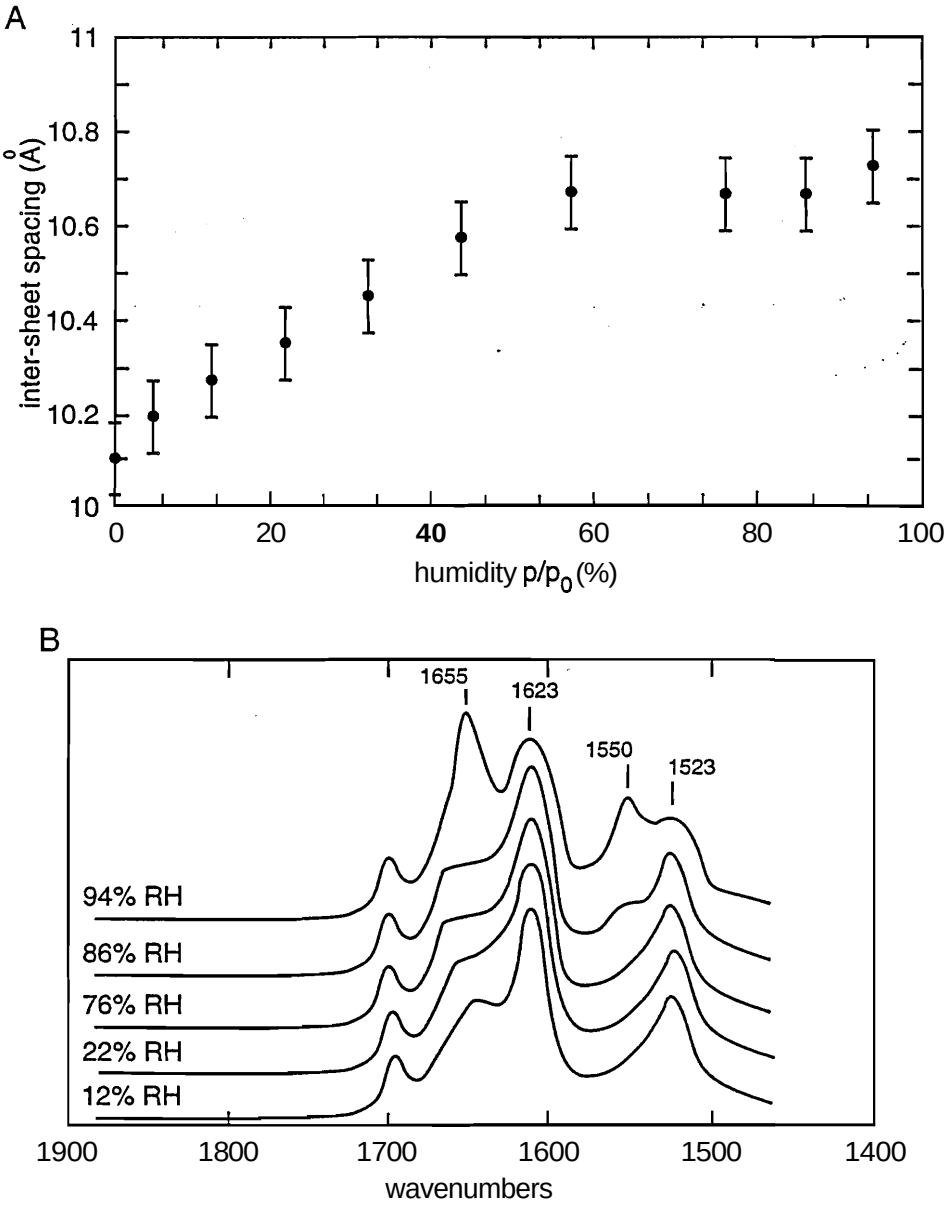


Figure 14-10. (A) Changes in intersheet spacing as a function of hydration. (B) Infrared amide I and II regions at different relative humidities. (C) Water content of hydrated samples in (A) and (B). Reprinted from Chen et al., 1995, with permission.

ties, and water contents of these same samples (from gravimetric measurements). The intersheet distance begins to change at low water contents, indicating hydration of the ionic groups at the lamellar surface; however, the infrared spectra reveal no changes in the chain conformation until the water content is quite high, that is, after the intersheet spacing has expanded, providing accessibility of the crystalline region to water.

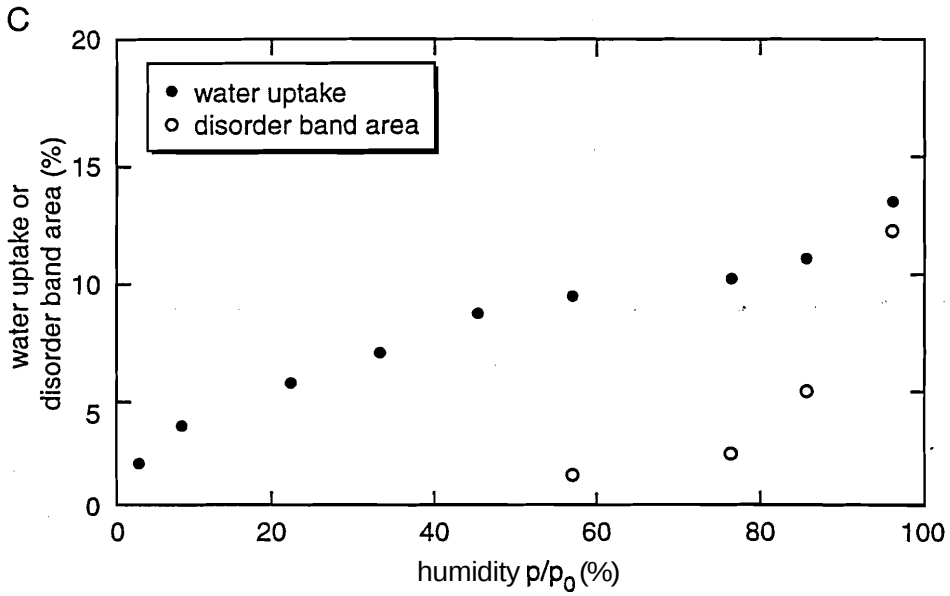


Figure 14-10. (Continued).

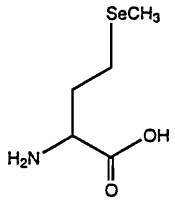
Artificial Proteins from Artificial Amino Acids

For some materials applications, it would be desirable to expand the set of monomer units available for construction of polymer chains beyond those normally used by organisms as substrates for protein synthesis. This section describes the incorporation of several non-natural amino acids—selenomethionine, *p*-fluorophenylalanine, 3-thienylalanine, 5′,5′,5′-trifluoroleucine, azetidine-2-carboxylic acid, and 3,4-dehydroproline—into genetically engineered artificial proteins expressed in bacterial hosts. Because of the relatively large sample size requirements of materials research, we have not yet exploited the intriguing approach of chemical misacylation of tRNAs and *in vitro* translation, a strategy developed and exploited successfully by others (Ellman et al., 1992; Hecht, 1992).

Selenomethionine

It has been known since the 1950s that selenomethionine (SeMet, 5) can be substituted for methionine in the *in vivo* synthesis of bacterial proteins (Cowie et al., 1959; Cowie & Cohen, 1957; Hendrickson et al., 1990; Tuve & Williams, 1957). SeMet supports growth of E. *coli* methionine auxotrophs (Cowie & Cohen, 1957; Tuve & Williams, 1957) and has been shown to replace methionine virtually completely in thioredoxin of E. *coli* and phage T4 (Hendrickson et al., 1990). In addition, SeMet, with the bulky selenide group in the side chain, would be expected to be excluded from the interior of lamellar crystals, and thereby provide a population of reactive functional groups at the lamellar surface (Dougherty et al., 1993). With these ideas in mind, we undertook the synthesis of polymers of repeating unit sequence -$(GlyAla)_3$GlySeMet-.

The expression plasmid chosen for these experiments (pGEX-2T) (Smith & Johnson, 1988) encodes a 26 kD fragment of glutathione-S-transferase (GST) at the amino terminus of the target protein. The host methionine auxotroph was prepared from E. *coli* strain HB101



5

F14-11

by mutagenesis with ethyl methanesulfonate and antibiotic selection in the absence of methionine (Miller, 1972). Auxotrophs were characterized by strict dependence on added methionine (or SeMet) for growth. As shown in Figure 14-11, substitution of SeMet for methionine allowed the mutant cells to grow, but with a two-fold increase in generation time.

To assess the level of SeMet substitution in the target protein, competition experiments were conducted using ^{35}S-methionine and unlabeled SeMet; labeled SeMet is not readily available. Cultures of transformed auxotrophs were grown to logarithmic phase in the presence of a small amount of methionine and with increasing amounts of SeMet.

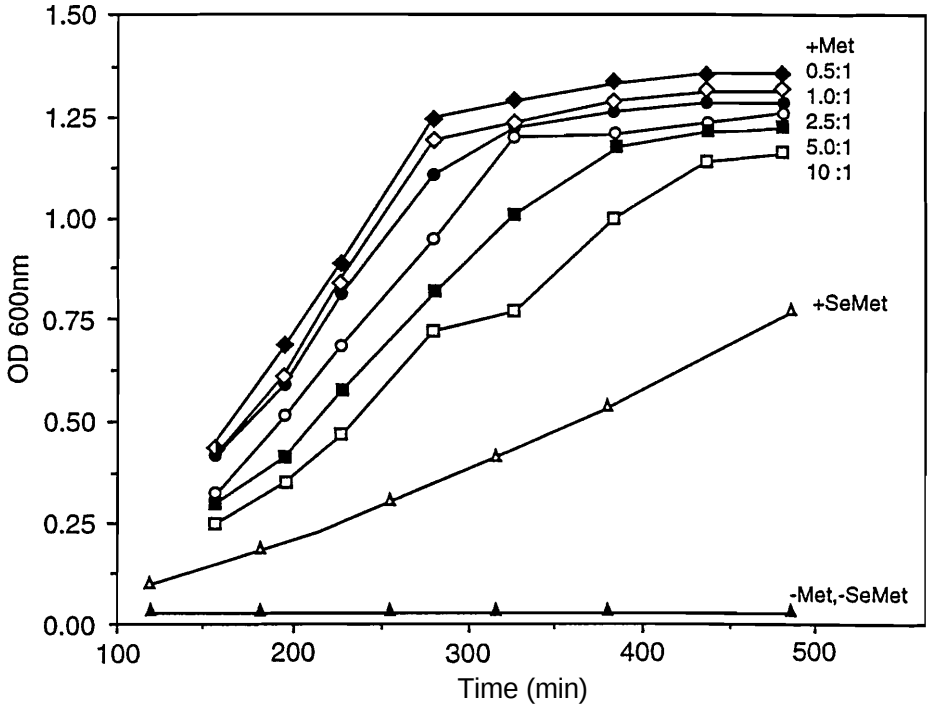


Figure 14-11. Growth kinetics of transformed E. *coli* cells grown in medium lacking methionine (−Met, −SeMet), with selenomethionine added without methionine (+SeMet), or methionine added with selenornethionine in different ratios (+Met). Reprinted from Dougherty et al., 1993, with permission.

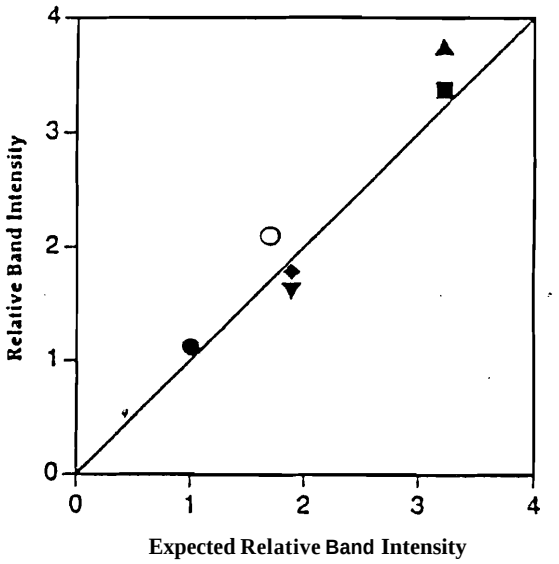
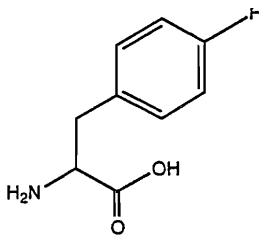


Figure 14-12. Effect of selenomethionine on the incorporation of ^{35}S-methionine into a polypeptide of repeating unit sequence -(GlyAla)$_3$GlySeMet-. Comparison of experimental and predicted incorporation values for the induced polypeptide at different ratios of selenomethionine to methionine. Experimental ratios were obtained by scanning autoradiographic signals. Each experimental value consists of intensity measurements taken from at least three separate time points and averaged to yield the experimental ratio. ●, 10:1/10X Met; ○, 2.5:1/5:1; ◆, 5:1/10X Met; ▼, 5:1/10:1; ■, 2:5:1/10:1; A, 2.5; 1/10X Met. 10X Met represents a control consisting of a 10-fold excess of unlabeled methionine. The theoretical ratios correspond to the intensities expected if the incorporation of selenomethionine is equivalent to that of methionine at each concentration tested. Reprinted from Dougherty et al., 1993, with permission.

^{35}S-Methionine was then added to the cultures, and protein synthesis was induced five minutes later by addition of isopropyl β-thiogalactopyranoside (IPTG). Samples were removed after varying periods of time, and analyzed by SDS polyacrylamide gel electrophoresis. Autoradiograms and stained gels were scanned densitometrically. Figure 14-12 illustrates the decrease in incorporation of radioactive methionine observed with increasing concentrations of SeMet in the growth medium; in fact, there is excellent agreement between the experimentally determined ratio of SeMet to methionine in the product, and the ratio of concentrations of the two amino acids in the medium. These data are consistent with complete or near complete substitution by SeMet.

p-Fluorophenylalanine

Polymers containing fluorinated amino acids are expected to exhibit many of the useful characteristics of conventional fluoropolymers, including low surface energy, low coefficient of friction, and excellent solvent resistance. In our first studies of genetically engineered fluoropolymers (Yoshikawa et al., 1994), p-fluorophenylalanine (pfF, *6)* was substituted for phenylalanine in a polymer comprising thirteen repeats of the octapeptide sequence -{(AlaGly)$_3$PheGly}-.



6

E. *coli* phenylalanine auxotrophs were generated by infecting a donor strain containing the *PheA*::Tn10 mutation with the transducing phage P1 and then transferring the transposon to the bacterial strain containing the expression vector pET3-b (Bochner et al., 1990; Miller, 1972; Rosenberg et al., 1987). This vector requires expression of bacteriophage T7 RNA polymerase for transcription of the target gene (Studier et al., 1990). In experiments directed toward incorporation of amino acid analogues, it might be anticipated that substitution of the analogue for the natural amino acid would result

OK as is

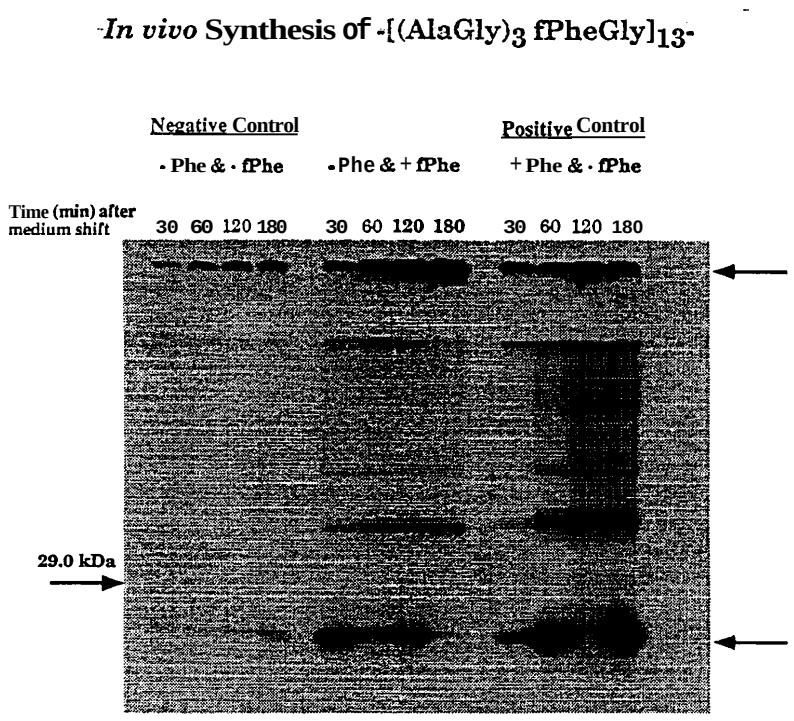


Figure 14-13. Expression of target protein labeled with ^{3}H-glycine. For lanes 1–4 the culture medium lacks both phenylalanine and p-fluorophenylalanine; for lanes 5–8, the medium contains p-fluorophenylalanine but lacks phenylalanine; for lanes 9–12, the medium contains phenylalanine but lacks p-fluorophenylalanine. The arrows indicate the target protein; time points are relative to medium shift. Reprinted from Yoshikawa et al., 1994, with permission.

121

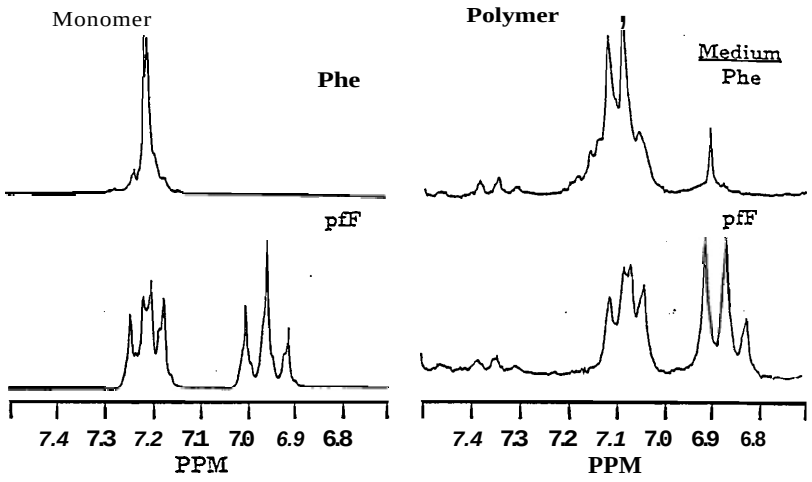


Figure 14-14. ^{1}H NMR spectra of -{(AlaGly)$_3$PheGly}- and -{(AlaGly)$_3$pfFGly}- in DCOOH. The aromatic regions of the spectra are compared with those for L-phenylalanine and p-fluoro-L-phenylalanine. The signal at 6.9 ppm for the protein containing phenylalanine is attributed to the solvent; this signal is obscured in the spectrum of protein containing p-fluorophenylalanine. Reprinted from Yoshikawa et al., 1994, with permission.

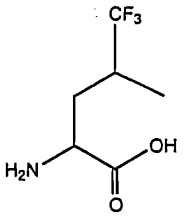
in an inactive polymerase. Therefore, we have adopted a protocol in which expression of the target protein begins with a ten minute induction in the presence of the twenty natural amino acids (but in the absence of pfF) followed by a shift to medium containing pfF after a pool of active polymerase has been established (Fig. 14-13).

Analysis of amino acid composition and integration of the relevant signals in the NMR spectra of target polymers made by such a protocol show a level of replacement of phenylalanine of 95 to 100% (Fig. 14-14, Table 14-3). The target proteins—{(AlaGly)$_3$PheGly} and {(AlaGly)$_3$pfFGly}—were also analyzed by Fourier transform

Table 14-3 Amino Acid Composition Analysis of Target Proteins Containing Phe and pff

Amino acid	Mol% (theoretical)	Mol% (observed for proteins containing pfF ~~Phe~~)	Mol% (observed for proteins containing ~~pfF~~ Phe)
Glycine	50.0	51.0	50.5
Alanine	37.7	37.0	36.4
Phenylalanine	12.3[a]	10.3	
p-Fluorophenylalanine	12.3[b]	0.6	11.2
Aspartic acid		1.1	1.0
Leucine			0.5
Valine			0.4

[a]Expected for proteins containing phenylalanine.
[b]Expected for proteins containing p-fluorophenylalanine.



7

infrared spectroscopy, which provides evidence for the antiparallel β-sheet architecture; the arnide I bands are observed at 1625 cm^{-1} and the amide II bands at 1525 cm^{-1} (Krimm & Abe, 1972; Moore & Krimm, 1976). Evidence for β-sheet structure is also observed in wide angle X-ray scattering patterns of oriented crystal mats.

5′,5′,5′-Trifluoroleucine

5′,5′,5′-Trifluoroleucine (TfL, 7) has been shown to be incorporated into proteins synthesized by E. *coli* leucine auxotrophs (Fenster & Anker, 1969), and incorporation of TfL into artificial proteins would be expected to allow placement of fluorinated residues at well-defined locations on the polymer chain.

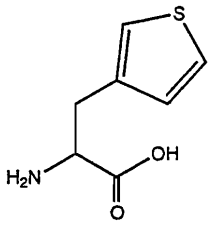
The sequence of the protein in which we investigated the substitution of TfL for leucine was $\{(GlyAla)_3GlyLeu\}_{12}GlyAla$ (Kothakota et al., unpublished). Gene synthesis, leucine auxotroph construction, and protein expression were carried out as described for expression of the protein containing pfF.

No target protein was produced by the host leucine auxotroph in the absence of leucine and TfL, while ^1H NMR spectra of proteins prepared in media supplemented with TfL indicated successful incorporation of the analogue. The level of substitution of TfL for leucine was estimated at a maximum of about 90% from NMR determination of the ratio of the methyl protons of TfL to those of alanine.

Of the stereoisomers of TfL (*2S,4S*; *2S,4R*; *2R,4R*; *2R,4S*), only the *2S* isomers, or L-amino acids, are used in biological protein synthesis. In ^{19}F NMR spectra of monomeric TfL and $\{(GlyAla)_3GlyTfL\}_{12}GlyAla$, two sets of signals, attributable to the *2S,4S*;*2R,4R* and *2S,4R*;*2R,4S* isomers, respectively, were observed both for the monomer and for the polymer, indicating that both of the 2*S* isomers are incorporated into protein to similar extents. The fact that either of the two diastereotopic methyl groups can be fluorinated without loss of translational activity, suggests that perhaps ***both*** can be fluorinated. Experiments with hexafluoroleucine are underway.

Infrared spectroscopy and X-ray diffraction studies indicate the presence of stacked antiparallel β-sheets in crystalline samples of both $\{(GlyAla)_3GlyLeu\}_{12}GlyAla$ and $\{(GlyAla)_3GlyTfL\}_{12}GlyAla$. Vibrational modes similar to those observed for the periodic proteins discussed above (and characteristic of antiparallel β-sheets) are observed in the IR spectra.

The surface properties of thin films of $\{(GlyAla)_3GlyLeu\}_{12}GlyAla$ and $\{(GlyAla)_3GlyTfL\}_{12}GlyAla$ were assessed by measuring the advancing contact angles for water and hexadecane. The fluorinated form of the polymer exhibits decreased wettability by hexadecane, a sensitive indicator of surface fluorination.



8

3-Thienylalanine

3-Thienylalanine (3-TA, **8)** was chosen for study because of its similarity to the 3-alkylthiophenes (Kothakota et al., 1995). Poly(3-alkylthiophene)s form excellent organic conductors, showing conductivities of about 2000 S cm^{-1} after doping (Roncali et al., 1988). The incorporation of 3-TA into engineered proteins may provide a means for electrodepositing such materials on electrodes or for fabrication of enzyme-based sensors or controlled delivery devices (Kothakota et al., 1995).

The target chosen for these experiments was a repeating polymer of sequence -{(GlyAla)$_3$GlyPhe}-$_{13}$, in which 3-TA was substituted for phenylalanine. Gel electrophoresis (Fig. 14-15) and amino acid composition analysis (Table 14-4) demonstrated efficient incorporation of 3-TA into recombinant proteins in place of phenylalanine.

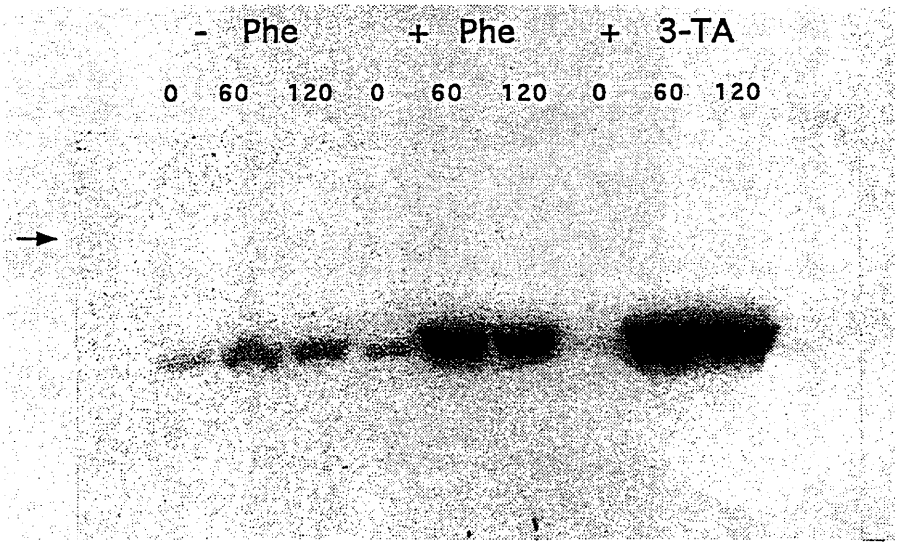


Figure 14-15. Autoradiogram of ^{3}H-labeled proteins produced by transformed phenylalanine auxotrophs induced to synthesize protein in the presence of the 20 natural amino acids and shifted after ten minutes to media lacking Phe and 3-TA, containing Phe, or containing 3-thienylalanine (3-TA). Reprinted from Kothakota et al., 1995, with permission.

Table 14-4 Amino Acid Compositions of Target Proteins Containing Phe or 3-Thienylalanine (3-TA)

Amino acid	Mol% (theoretical)	Mol% (observed for -{(GlyAla)$_3$GlyPhe}-$_{13}$	Mol% (observed for -{(GlyAla)$_3$Gly$_3$-TA}-$_{13}$
Glycine	50.0	49.4	47.0
Alanine	37.7	35.6	34.1
Phenylalanine	12.3[a]	11.7	2.2
3-Thienylalanine	12.3[b]		10.8

[a]Expected for -{(GlyAla)$_3$GlyPhe}-$_{13}$.
[b]Expected for -{(GlyAla)$_3$Gly$_3$-TA}-$_{13}$.

F14-16

The relative amounts of phenylalanine and 3-TA in the product were also assessed quantitatively by ultraviolet spectroscopy (which takes advantage of the fact that the phenylalanine absorption maximum lies at 256 nm while that for 3-TA is at 233 nm), and by NMR spectroscopy. Figure 14-16 shows UV spectra for target proteins produced by cultures in which protein synthesis was induced in media supplemented either with phenylalanine or with 3-TA; the phenylalanine absorption maximum is absent from the spectrum of the target protein containing the electroactive analogue. Three sets of signals arising from the monosubstituted thiophene ring are observed in the NMR spectrum of -{(GlyAla)$_3$Gly3-TA}-$_{13}$, while two broad resonances attributed to the phenyl protons appear in the spectrum of -{(GlyAla)$_3$GlyPhe}-$_{13}$. Integration of the

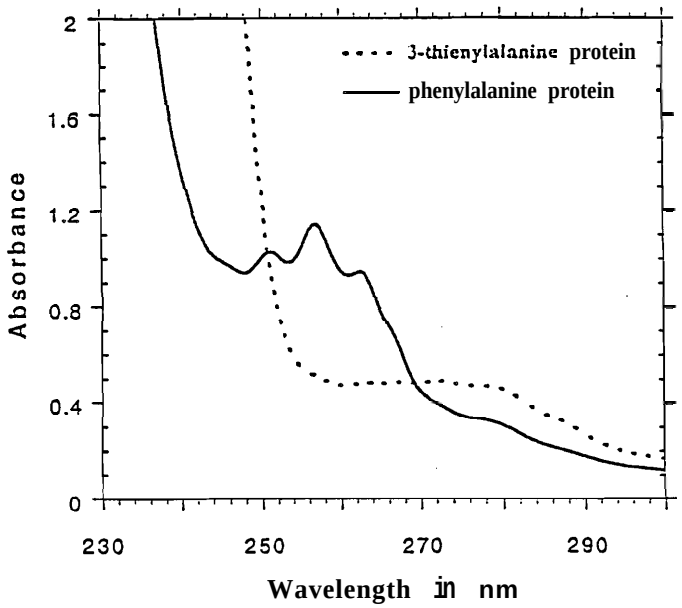


Figure 14-16. UV spectra of -{(GlyAla)$_3$GlyPhe}-$_{13}$ and -{(GlyAla)$_3$Gly3-TA}-$_{13}$. Reprinted from Kothakota et al., 1995, with permission.

spectrum of the 3-TA polymer indicates the extent of substitution of 3-TA for phenylalanine to be approximately 80%. Artificial proteins containing 3-TA have now been fabricated into composite films with poly(3-methylthiophene); studies of the electrochemical properties of such films are in progress.

Azetidine-2-carboxylic Acid

Earlier in this chapter, we showed that repeating polypeptides of the sequence -(AlaGly)$_n$ProGluGly- adopted disordered structures in the solid state. To explore the role of proline in promoting (or frustrating) the formation of β-sheets, azetidine-2-carboxylic acid (Aze, **9**) was substituted for proline in polypeptides of repeating sequence -(AlaGly)$_3$ProGluGly- (Deming et al., 1996). Previous studies have shown that the effects of incorporating azetidine-2-carboxylicacid into proteins can be striking (Mauger & Witkop, 1966); replacement of about 4% of the proline in collagen with Aze destabilizes the triple helix bundle (Lane et al., 1971). Computational studies have suggested that proteins containing Aze are more flexible than those containing proline, and more likely to adopt β-turn geometries (Zagari et al., 1990). Both characteristics should contribute to assembly of ordered β-sheets in periodic polypeptides.

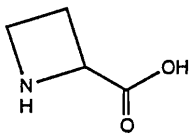
Because Aze does not support bacterial growth (Peterson & Fowden, 1963), we grew transformed cells in medium supplemented with proline and then induced synthesis of the target protein in medium containing Aze. Autoradiograms of ^{3}H-labeled proteins produced in transformed proline auxotrophs show that more protein is produced in Aze-supplemented growth medium than in medium containing neither proline nor Aze. Direct determination of the relative amounts of proline and Aze in the product by amino acid analysis is precluded by the fact that Aze is degraded under the conditions used to hydrolyze and analyze peptides (Fowden, 1956). Nevertheless, the Aze content can be estimated reliably at about 25%, by determination of the extent of diminution of proline concentration (Table 14-5).

T14-5

Also, ^{1}H NMR spectroscopy can be used to measure the level of Aze incorporation; the intensities of the proline resonances (3.62 ppm), compared to those of the alanine and glycine resonances (1.32 and 4.20 ppm, respectively) indicate that Aze is incorporated to a level of about 40%, somewhat higher than the estimate based on amino acid analysis.

The Pro and Aze forms of the polymers exhibit marked differences in pH-dependent solubility. Both polymers precipitate when the glutamate side chains are acid titrated, but partial replacement of proline with Aze results in a shift of the cloud point from pH 5.5 for the -(AlaGly)$_3$ProGluGly- polymers to pH 6.5 for those containing Aze.

Beta-sheet structure was indeed observed in -(AlaGly)$_3$ProGluGly- polymers containing Aze as shown by the typical amide vibrational modes in the infrared spectrum



9

Table 14-5 Amino Acid Compositions of -(AlaGly)$_3$ProGluGly-Repeating Proteins Made in Media Supplemented with Pro or Aze

Amino Acid	Mol% (theoretical)	Mol% (observed)[a]	Mol% (observed)[b]
Glycine	34.5	35.3	33.8
Alanine	29.0	27.9	28.1
Glutamic acid	9.5	10.6	10.9
Proline	10.0[c]	10.1	7.4

[a]Media supplemented with Pro.
[b]Media supplemented with Aze.
[c]Expected for proteins made in medium containing proline.

F14-17

(Fig. 14-17). The Aze substituted polymer exhibited amide I and amide II absorptions at 1629 and 1522 cm^{-1}, respectively, as expected for proteins containing β-sheet structures (Moore & Krimm, 1976) while the corresponding vibrations observed for the proline form of the polymer (1654 and 1534 cm^{-1}) are not characteristic of p-sheets, as discussed previously (Bandekar & Krimm, 1986).

The changes observed in the structure of -(AlaGly)$_3$ProGluGly- upon low level substitution by Aze are remarkable, in view of the fact that only four to six residues, out of a total of 148 in the polymer chain, are modified. The factors giving rise to the conformational ordering of Aze-substituted polymers are currently under study.

3,4-Dehydroproline

3,4-Dehydroproline (Dhp, 10) provides a means of introducing alkene groups into polypeptide chains at precisely determined locations. These groups are then available for subsequent site-specific chemical modifications, including oxidations, halogenations, and so forth.

Dhp is incorporated into protein much more efficiently than Aze (Mauger & Witkop, 1966; Rosenbloom & Prockop, 1970; Smith et al., 1962; Tristram & Thurston, 1966). In our experience, analysis of amino acid composition and ^1H NMR spectra gives levels of incorporation ranging from 90 to 100% (Deming et al., unpublished results).

Dhp in monomer form can be treated with H_2O_2 or with Br_2 in formic acid to produce 3,4-dihydroxyproline or 3,4-dibromoproline (Buku et d., 1980; Nakajima & Volcani, 1969), respectively, in virtually quantitative yields. Similar reactions can be carried out on Dhp polymers with essentially complete conversion of the Dhp residues, as indicated by the disappearance of the alkene proton resonances at 6.12 and 5.88 ppm in the ^1H NMR spectra.

Uniform Helical Rodlike Polymers

Poly(α,L-glutamic acid) (PLGA) and its derivatives, as produced by traditional synthetic methods, are characterized by the same kinds of broad molecular weight distributions that one finds in synthetic polymers generally (Block, 1983). In addition, PLGA

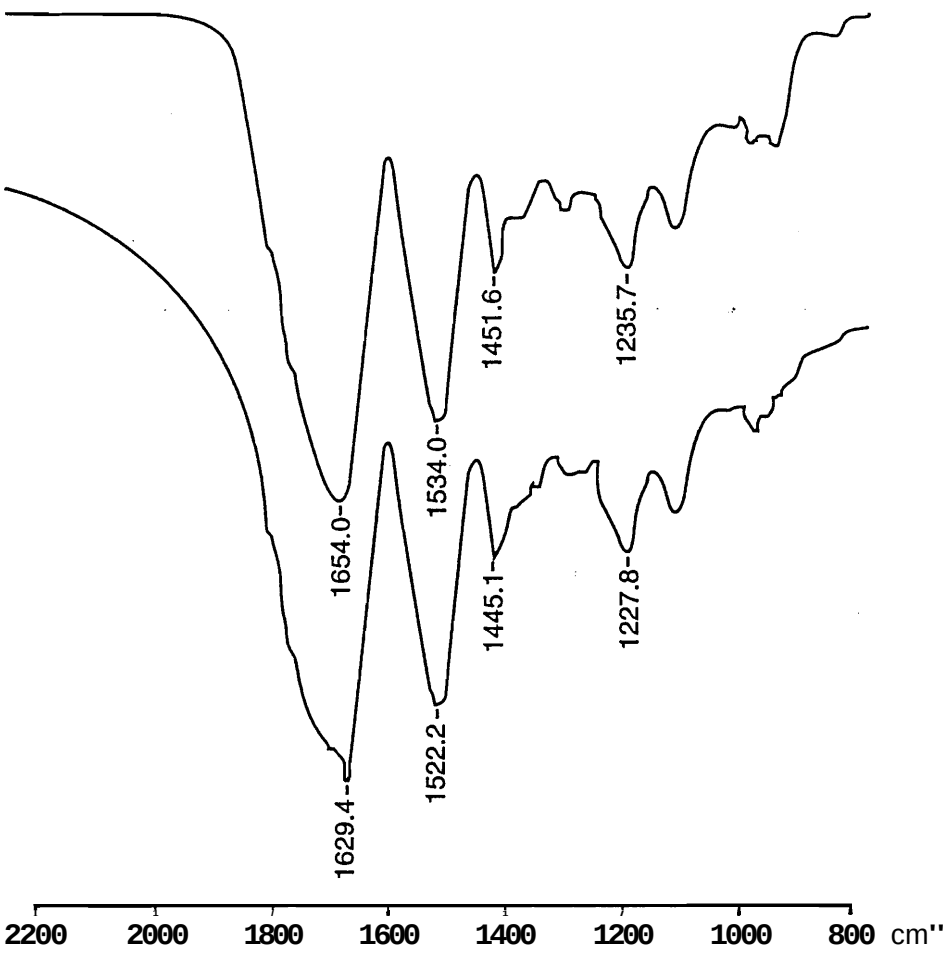
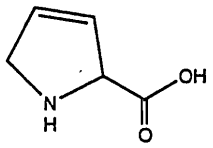


Figure 14-17. FTIR spectroscopy of unsubstituted -(AlaGly)$_3$ProGluGly- polymer (top) and -(AlaGly)$_3$ProGluGly- polymer containing Aze (bottom). Reprinted from Deming et al., 1996, with permission.

is usually synthesized in the ester form (e.g., as the benzyl ester, obtained by ring-opening polymerization of the corresponding N-carboxy-a-amino acid anhydride) and subsequent reactions that convert the ester to the parent acid can result in substantial racemization (Block, 1983). PLGA has been used in fundamental studies of polyelec-



10

trolytes and of the helix-coil transition (Poland & Scheraga, 1970), and poly(γ-benzyl α,L-glutamate) (PBLG), the most widely studied ester of PLGA, forms liquid crystalline solutions (Horton et al., 1990) and ordered monolayer films (McMaster et al., 1991). Because the heterogeneity of these materials complicates studies of their physical properties, production of monodisperse PLGA and various derivatives should provide significant new insights into the behavior of helical, rodlike macromolecules. Our own interest is directed in large part toward the development of novel smectic mesophases based on monodisperse PBLG.

We have used the biosynthetic strategy described in previous sections to prepare monodisperse derivatives of PLGA (Zhang et al., 1992). An oligonucleotide sequence encoding seventeen glutamic acid residues followed by a single aspartic acid unit, was synthesized using phosphoramidite chemistry (McBride & Caruthers, 1983). Codons for aspartic acid were included in the coding sequence to provide recognition and cleavage sites for the restriction endonuclease, *BbsI*, which was used to isolate the coding sequence after amplification. Although it is possible to devise coding sequences that use only glutamic acid codons, incorporation of flanking restriction sites into such a sequence would require that a single codon (either GAA or GAG) be used to construct the remainder of the oligonucleotide. We preferred to avoid such a design because of concern about genetic instability.

A population of multimers of the oligonucleotide sequence was produced in E. *coli* strain DH5αF′, and a tetramer of this sequence was isolated and inserted into the expression vector, pGEX-3X, which yields the target protein as a fusion with a 26 kD fragment of glutathione-S-transferase at its amino terminus. The protein was purified by affinity chromatography on a glutathione-linked matrix, and subjected to CNBr cleavage to remove the GST fragment. Analysis by electrophoresis under nondenaturing conditions revealed a single tight protein band. Comparison of PLGA produced via the biosynthetic route, with commercially available samples of PLGA produced by conventional chemical synthesis (Zhang et al., 1992), is striking and serves as a powerful example of the precision of the biosynthetic strategy in the production of monodisperse polymeric materials. Studies of the benzylation of PLGA, and of the assembly behavior of the resulting poly(γ-benzyl α,L-glutamate)s, are underway.

Hybrid Artificial Proteins: Combining Materials Properties with Biological Function

Naturally occurring proteins often have two or more domains that are responsible for different activities such as recognition, binding, and catalysis (Baneyx et al., 1990; Hunger et al., 1990). Using the biosynthetic approach described in this chapter, it is possible to extend this idea to the construction of hybrid artificial proteins that combine the materials properties of synthetic polymers and the biological functions of natural proteins.

We have expressed hybrid proteins composed of alanylglycine-richleader sequences ({(AlaGly)$_3$GluGly}$_{36}$ (Dong et al., 1994) or {(AlaGly)$_3$ProGluGly}$_{16}$) (Dong et al., 1996) attached at the N-terminus of the phosphotriesterase of Pseudomonas diminuta, which catalyzes efficient hydrolysis of organophosphorus compounds used as nerve agents and pesticides (Dumas & Raushel, 1989). The function of the artificial domains

Table 14-6 Catalytic Properties of Hybrid Protein and Phosphotriesterase

Parameter	Hybrid Protein	Phosphotriesterase
Specific activity (units/mg protein)	1976	7434
Michaelis constant (μM)	106	155
Catalytic rate constant (s^{-1})	2315	6540

[a]One unit of enzyme activity will hydrolyze 1 mmol of paraoxon per minute.

in these hybrids is to provide a simple and efficient means of immobilizing the phosphotriesterase on glass or on other basic surfaces. "Self-immobilizing" enzymes of this type might be expected to find application in detection of pesticides and in detoxification of nerve agents.

Polynucleotides encoding the artificial domains were constructed as described previously and ligated at the 5'-end of a plasmid-borne DNA sequence encoding the phosphotriesterase. The recombinant genes were expressed in E. *coli* and the hybrid proteins were purified by metal chelate affinity chromatography, taking advantage of an oligohistidine sequence encoded at the amino terminus.

Enzyme activity was detected in whole cell lysates and in soluble fractions subjected to metal chelate affinity chromatography, by monitoring the hydrolysis of the pesticide paraoxon to diethyl phosphate and the p-nitrophenolate anion, which absorbs strongly at 400 nm.

When cell lysates or purified proteins are loaded onto a DEAE-Sephadex anion-exchange column, the hybrid protein binds to the resin via the acidic, repetitive polypeptide domain. If paraoxon is then added to the column, the characteristic yellow color of the p-nitrophenolate anion is observed. Similar treatment of the unmodified phosphotriesterase reveals no activity on the column, as the enzyme has no strong affinity for the cationic resin. This experiment illustrates the functions of each of the two domains in the hybrid protein: binding activity arising from the artificial leader sequence, and catalytic activity from the phosphotriesterase domain.

The catalytic activity of the hybrid protein carrying the $\{(AlaGly)_3ProGluGly\}_{16}$ leader sequence is compared to that of unmodified phosphotriesterase in Table 14-6. While the specific activity and the catalytic rate constant of the hybrid protein are lower than those of the unmodified enzyme, the hybrid protein also has a lower Michaelis constant, indicating a higher affinity for substrate than that of the free phosphotriesterase. The pH profiles of catalytic activity of the hybrid protein and the unmodified phosphotriesterase are also very similar in the range of pH 6 to pH 11. Previous studies have shown that the bacterial phosphotriesterase has a pK_a of 6.1, indicating, together with chemical modification studies, that catalytic activity depends on the ionization of a histidine residue (Dumas & Rauschel, 1990). We are currently exploring the immobilization of hybrid phosphotriesterases onto optical fibers for use as biosensors.

Summary

Protein engineering is opening important new opportunities in the design and synthesis of new kinds of polymeric materials and materials systems. By using a protein engineering approach, one can prepare new macromolecules with precise control of chain length, sequence and stereochemistry, and thereby control conformational properties and assembly behavior. Progress has been made in the engineering of macromolecular crystals (the "solid-state protein-folding problem"), and the prospects appear excellent for the design of new kinds of liquid crystal phases and surface arrays. The capacity of bacterial cells to accommodate nonnatural amino acids as substrates for protein biosynthesis extends the scope of the method, and has already led to artificial proteins with unusual surface properties and electrochemical behavior. Finally, the combination of artificial and natural protein domains provides a route to "hybrid proteins" in which both materials properties and biological function are subject to control.

Volkin, D. B., Staubli, A., Langer, R., & Klibanov, A. M. (1991) *Biotechnol. Bioeng.* 37, 843–853.
Waldmann, H. (1988) *Tetrahedron Lett.* 29, 1131–1134.
Wallace, J. S. & Morrow, C. J. (1989) *J. Polym. Sci. Part A: Polym. Chem.* 27, 2553–2567.
Wandrey, C. (1986) in *Enzymes as Catalysts in Organic Synthesis* (Schneider, M. P. & Reidel, D., Eds.) pp 263, Dordrecht, Holland.
Wehtje, E., Adlercreutz, P., & Mattiasson, B. (1993) *Biotechnol. Bioeng.* 41, 171–178.
Wescott, C. R. & Klibanov, A. M. (1993) *J. Am. Chem. Soc.* 115, 1629–1631.
Whitesides, G. M. & Wong, C.-H. (1985) *Angew. Chem. Int. Ed. Eng.* 24, 617–638.
Wirth, P., Souppe, J., Tritsch, D., & Biellmann, J.-F. (1991) *Bioorg. Chem.* 19, 133–142.
Wong, C.-H. (1992) *Trends Biotechnol.* 10, 337–341.
Wong, C.-H., Ichikawa, Y., Krach, T., Narvor, C. G., Dumas, D. P., & Look, G. C. (1991a) *J. Am. Chem. Soc.* 113,8137–8145.
Wong, C.-H., Liu, K. K.-C., Kajimoto, T., Chen, L., Zhong, Z., Dumas, D. P., Liu, J. L.-C., Ichikawa, Y., & Shen, G.-J. (1992) *Pure Appl. Chem. 64,* 1197–1202.
Wong, C.-H., Shen, G.-J., Pederson, R. L., Wang, Y.-F., & Hennen, W. J. (1991b) *Methods Enzymol. 202,* 591–620.
Yamazaki, Y. & Hosono, K. (1988) *Tetrahedron Lett.* 29,5769–5770.
Yang, F. X. & Russell, A. J. (1993) *Biotechnol. Prog.* 9,234–241.
Yang, Z. & Robb, D. A. (1991) *Biochem. Soc. Trans.* 20, 13S.
Yang, Z. & Robb, D. A. (1994) *Biotechnol. Bioeng.* 43, 365–370..
Yang, Z., Robb, D. A., & Halling, P. J. (1992) in *Biocatalysis in Non-conventional Media* (Tramper, J., Vermüe, M. H., Beeftink, H. H., & von Stockar, U., Eds.) pp 585–592, Elsevier, Amsterdam.
Yang, Z., Zacherl, D., & Russell, A. J. (1994) *J. Am. Chem. Soc. 115,* 12251–12257.
Yoshimoto, T., Takahashi, K., Nishimura, H., Ajima, A., Tamaura, Y., & Inada, Y. (1984) *Biotechnol. Lett.* 6, 337–340.
Yoshimoto, T., Ritani, A., Ohwada, K., Takahashi, K., Kodera, Y., Matsushima, A., Saito, Y., & Inada, Y. (1987) *Biochem. Biophys. Res. Commun.* 148, 876882.
Zaks, A. & Klibanov, A. M. (1984) *Science* 224, 1249–1251.
Zaks, A. & Klibanov, A. M. (1985) *Proc. Natl. Acad. Sci. U.S.A.* 82,3192–3196.
Zaks, A. & Klibanov, A. M. (1986) *J. Am. Chem. Soc. 108,* 2767–2768.
Zaks, A. & Klibanov, A. M. (1988a) *J. Biol. Chem.* 263,3194–3201.
Zaks, A. & Klibanov, A. M. (1988b) *J. Biol. Chem. 263,* 8017–8021.

Chapter 14

Aota, S., Gojobori, T., Ishibashi, F., Maruyama, T, & Ikemura, T. (1988) *Nucleic Acids Res.* 16, R315–R402.
Atkins, E. D. T., Hill, M., Hong, S. K., Keller, A., & Organ, S. (1992) *Macromolecules 25,* 917–924.
Bahr, U., Deppe, A., Karas, M., Hillenkamp, F., & Giessmann, U. (1992) *Anal. Chem.* 64, 2866–2869.
Bandekar, J. and Krimm, S. (1986) in *Advances in Protein Chemistry* (Anfinsen, C. B., Edsall, J. T., Richards, F. M., Eds.) Vol. 38, pp 183–364, Academic Press, Orlando, FL.
Baneyx, F., Schmidt, C., & Georgiou, G. (1990) *Enzyme Microb. Technol.* 12, 337–342.
Beavis, R. C., Chait, B. T., Creel, H. S., Fournier, M. J. Mason, T. L., & Tirrell, D.A. (1992) *J. Am. Chem. Soc.* 114,7584–7585.
Block H. (1983) *Poly(γ-benzyl-L-glutamate) and Other Glutamic Acid Containing Polymers,* Gordon and Breach, New York.
Bochner, B. R., Huang, H., Schieven, G. L., & Ames, B. N. (1990) *J. Bacteriol. 143,* 926.
Brown, *L.* & Trotter, I. F. (1956) *Trans. Faraday Soc.* 52,537–548.

Buku, A., Faulstich, H., Wieland, T., & Dabrowski, J. (1980) ***Proc. Natl. Acad. Sci. U.S.A.*** 77, 2370–2371.
Cappello, J., Crissman, J., Dorman, M., Mikolajczak, M., Textor, G., Marquet, M., & Ferrari, F. (1990) ***Biotechnol. Prog.*** 6, 198–202.
Chen, C. C., Krejchi, M. T., Tirrell, D. A., & Hsu, S. L. (1995) ***Macromolecules*** 28, 1464–1469.
Chou, P. Y. & Fasman, G. D. (1974a) ***Biochemistry 13***, 211–222.
Chou, P. Y. & Fasman, G. D. (1974b) ***Biochemistry 13***, 222–245.
Cowie, D. B. & Cohen, G. N. (1957) ***Biochim. Biophys. Acta*** 26,252–261.
Cowie, D. B., Cohen, G. N., Bolton, E. T., & De Robichon-Szulmajster, H. (1959) ***Biochim. Biophys. Acta*** 34, 39–46.
Deguchi, Y., Fournier, M. J., Mason, T. L., & Tirrell, D. A. (1994) ***J. Macromol. Sci., -Pure Appl. Chem. A31***, 1691–1700.
Deming, T. J., Fournier, M. J., Mason, T. L., & Tirrell, D. A. unpublished results.
Deming, T. J., Foumier, M. J., Mason, T. L., & Tirrell, D. A. (1996) ***Macromolecules*** 29, 1442–1444.
Dong, W., Fournier, M. J., Mason, T. L., & Tirrell, D. A. (1994) ***Polymer Preprints*** 35,419.
Dong, W., Foumier, M. J., Mason, T. L., & Tirrell, D. A. (1996) ***Polym. Mat. Sci. Eng.*** 74, 71–72.
Dougherty, M. J., Kothakota, S., Mason, T. L., Tirrell, D. A., & Foumier, M. J. (1993) ***Macromolecules*** 26, 1779–1781.
Dreyfuss, P. & Keller, A. (1970a) ***J. Macromol. Sci.-Phys. B4***, 811–835.
Dreyfuss, P. & Keller, A. (1970b) ***J. Polym. Sci. Part B., Polym. Phys. 8***, 253–258.
Dumas, D. P. & Raushel, F. M. (1989) ***J. Biol. Chem.*** 264, 19659–19665.
Dumas, D. P. & Rauschel, F. M. (1990) ***J. Biol. Chem.*** 265, 21498–1503.
Ellman, J. A., Mendel, D., & Schultz, P. G. (1992) ***Science*** 255, 197–200.
Fenster, E. D. & Anker, H. S. (1969) ***Biochemistry 8***, 269–274.
Fowden, L. (1956) ***Biochem. J.*** 64,323–332.
Fraser, R. D. B. MacRae, T. P., Stewart, H. C., & Suzuki, E. (1965) ***J. Mol. Biol. 11***, 706–712.
Frushour, B. G. & Koenig, J. L. (1975) ***Biopolymers 14***, 2115–2135.
Geddes, A. J., Parker, K. D., Atkins, E. D. T., & Beighton, E. (1968) ***J. Mol. Biol.*** 32,343–358.
Hecht, S. M. (1992) ***Acc. Chem. Res.*** 25, 545–552.
Hendrickson, W. A., Horton, J. R., & LeMaster, D. M. (1990) ***EMBO J.*** 9, 1665–1672.
Hirichsen, G. (1973) ***Makromol. Chem. 166***, 291.
Horton, J. C., Donald, A. M., & Hill, A. (1990) ***Nature 346***, 44–45.
Hunger, H. D., Flachmeier, C., Schmidt, C., Behrendt, G., & Coutelle, C. (1990) ***Anal. Biochem. 187***, 89–93.
Ishida, M., Asakura, T., Yokoi, M., & Saito, H. (1990) ***Macromolecules*** 23, 88–94.
Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M., & Hecht, M. H. (1993) ***Science*** 262, 1680–1685.
Keller, A. (1957) ***Philos. Mag.*** 2, 1171–1175.
Keller A. (1959) ***J. Polym. Sci.*** 36,361–382.
Kothakota, S., Mason, T. L., Tirrell, D. A., Atkins, E. D. T., & Fournier, M. J., unpublished results.
Kothakota, S., Mason, T. L., Tirrell, D. A., & Fournier, M. J. (1995) ***J. Am. Chem. Soc.*** 117, 536–537.
Krejchi, M. T., Atkins, E. D. T., Waddon, A. J., Fournier, M. J., Mason, T. L., & Tirrell D. A. (1994) ***Science*** 265, 1427–1432.
Krejchi, M. T., Cooper, S. J. Deguchi, Y., Atkins, E. D. T. Fournier, M. J., Mason, T. L., & Tirrell, D. A. (1997) ***Macromolecules*** 30, 5012–5024.
Krimm, S. & Abe, Y. (1972) ***Proc. Natl. Acad. Sci. U.S.A.*** 69, 2788–2792.
Krimm, S. & Bandekar, J. (1986) ***Adv. Protein Chem.*** 38, 181–364.
Lane, J. M., Dehm, P., & Prockop, D. J. (1971) ***Biochim. Biophys. Acta*** 236, 517–527.

Lucas, F., Shaw, T. B., & Smith, S. G. (1960) J. Mol. Biol. 2, 339–349.
Mauger, A. B. & Witkop, B. (1966) Chem. Rev. *66*, 47–86.
McBride, L. J. & Caruthers, M. H. (1983) Tetrahedron Lett. 24, 245–248.
McGrath, K. P., Fournier, M. J., Mason, T. L., & Tirrell, D. A. (1992) J. Am. Chem. *Soc*. 114, 727–733.
McMaster, T. C., Carr, H. J., Miles, M. J., Cairns, P., & Morris, V. J. (1991) Macromolecules 24, 1428–1430.
Miller, J. H. (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, NY.
Miyazawa, T. & Blout, E. R. (1961) J. Am. Chem. *Soc*. 83,712–719.
Moore, W. H. & Krimm, S. (1976) Biopolymers *15*, 2465–2483.
Murthy, N. S., Stamm, M., Sibilia, J. P., & Krimm, S. (1989) Macromolecules 22, 1261–1267.
Nakajima, T., & Volcani, B. E. (1969) Science 164, 1400–1401.
Nemethy, G. & Printz, M. P. (1972) Macromolecules 5, 755–761.
Parkhe, A. D., Fournier, M. J., Mason, T. L., & Tirrell, D. A. (1993) Macromolecules 26, 6691–6693.
Peterson, P. J. & Fowden, L. (1963) Nature 200, 148–151.
Poland, D. & Scheraga, H. A. (1970) Theory *of Helix-Coil* Transitions in Biopolymers, Academic Press, New York.
Roncali, J., Yassar, A., & Garnier, F. (1988) J. Chem. *Soc*., Chem. Commun., 581–582.
Rosenberg, A. H., Lade, B. N., Chui, D., Lin, S., Dunn, J. J., & Studier, F. W. (1987) Gene 56, 125–135.
Rosenbloom, J. & Prockop, D. J. (1970) J. Biol. Chem. *245*, 3361–3368.
Saito, H., Tabeta, R., Asakura, T., Iwanaga, Y. Shoji, A., Ozaki, T., & Ando, I. (1984) Macromolecules 17, 1405–1412.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Second Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Smith, D. B. & Johnson, K. S. (1988) Gene *67*, 31–40.
Smith, L. C., Ravel, J. M., Skinner, C. G., & Shive, W. (1962) Arch. Biochem. Biophys. 99, 60-64.
Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990) Methods *Enzymol*. *185*, 60–89.
Tirrell, D., Grossman, S., & Vogl, O. (1979) Makromol. Chem. 180, 721–736.
Tristram, H. & Thurston, C. F. (1966) Nature *212*, 74–75.
Tuve, T. & Williams, H. (1957) J. Am. Chem. *Soc*. 79,5830–5831.
Vergalati, C., Imberty, A., & Perez S. (1993) Macromolecules 26, 4420–4425.
Warwicker, J. O. (1960) J. Mol. Biol. 2, 350–362.
Yoshikawa, E., Fournier, M. J., Mason, T. L., & Tirrell, D. A. (1994) Macromolecules 27, 5471–5475.
Zagari, A., Nemethy, G., & Scheraga, H. A. (1990) Biopolymers 30,951–959.
Zhang, G., Fournier, M. J., Mason, T. L., & Tirrell, D. A. (1992) Macromolecules 25,3601–3603